applied sciences

MDPI

Article

WaQuPs: A ROS-Integrated Ensemble Learning Model for Precise Water Quality Prediction

Firna Firdiani [1], Satria Mandala [1,*], Adiwijaya [1] and Abdul Hanan Abdullah [2]

[1] Human Centric Engineering & School of Computing, Telkom University, Jl. Telekomunikasi No. 1, Bandung 40257, West Java, Indonesia; vyrnvyrd@gmail.com (F.F.); adiwijaya@telkomuniversity.ac.id (A.)
[2] Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia; hanan@utm.my
* Correspondence: satriamandala@telkomuniversity.ac.id

Abstract: Water presents challenges in swiftly and accurately assessing its quality due to its intricate composition, diverse sources, and the emergence of new pollutants. Current research tends to oversimplify water quality, categorizing it as potable or not, despite its complexity. To address this, we developed a water quality prediction system (WaQuPs), a sophisticated solution tackling the intricacies of water quality assessment. WaQuPs employs advanced machine learning, including an ensemble learning model, categorizing water quality into nuanced levels: potable, lightly polluted, moderately polluted, and heavily polluted. To ensure rapid and precise dissemination of information, WaQuPs integrates an Internet of Things (IoT)-based communication protocol for the efficient delivery of detected water quality results. In its development, we utilized advanced techniques, such as random oversampling (ROS) for dataset balance. We used a correlation coefficient to select relevant features for the ensemble learning algorithm based on the Random Forest algorithm. Further enhancements were made through hyperparameter tuning to improve the prediction accuracy. WaQuPs exhibited impressive metrics, achieving an accuracy of 83%, precision of 82%, recall of 83%, and an F1-score of 82%. Comparative analysis revealed that WaQuPs with the Random Forest model outperformed both the XGBoost and CatBoost models, confirming its superiority in predicting water quality.

Keywords: Internet of Things; machine learning; water quality prediction; Random Forest; random oversampling

Citation: Firdiani, F.; Mandala, S.; Adiwijaya; Abdullah, A.H. WaQuPs: A ROS-Integrated Ensemble Learning Model for Precise Water Quality Prediction. *Appl. Sci.* **2024**, *14*, 262. https://doi.org/10.3390/app14010262

Academic Editor: Douglas O'Shaughnessy

Received: 14 November 2023
Revised: 21 December 2023
Accepted: 22 December 2023
Published: 28 December 2023

1. Introduction

Water is an irreplaceable natural resource, serving as a foundational element for the human body and playing an indispensable role in our survival. Beyond sustaining bodily functions, water is integral to a myriad of daily activities, including cooking, washing, and bathing. Nevertheless, swiftly and accurately assessing water quality poses a significant challenge due to its intricate composition, diverse sources, and the introduction of new pollutants. The task of promptly and precisely evaluating processed water quality remains an enduring challenge yet to be fully resolved. To address this issue, one promising avenue involves the application of machine learning techniques.

Numerous studies have explored the application of machine learning in the classification and prediction of water quality. Iyer et al. [1] conducted research on predicting water quality using machine learning, employing SVM, Random Forest, and Decision Tree models. The findings of their study revealed that the performance of the Random Forest model surpassed the other models, achieving an accuracy rate of 68%. However, it is noteworthy that the accuracy rate obtained by Random Forest in this study remained below the threshold of 70%.

Continuing the exploration of water quality prediction, Sen et al. [2] conducted further research focusing on addressing challenges in aquaculture through the introduction of an